

This is draft copy article:

1. Nedzved A., Starovoitov V. Nuclear Segmentation for high magnification of histology samples / Computational Vision and Medical Image Processing VipIMAGE 2009

## Nuclear Segmentation for high magnification of histology samples

A. Nedzved & V. Starovoitov

*United Institute of Informatics Problems of the National Academy of Sciences, Minsk, Belarus*

**ABSTRACT:** Quantitative measurement of nuclei in a histological section is a basic part in cancer diagnostics. Digital images of nucleuses registered from histological samples with high optical magnification are usually very noisy and require sophisticated preprocessing. We present a general method of histological sample processing for nuclear analysis and an algorithm for a nuclear contour detection. An algorithm for a nuclear structure extraction is also proposed in the paper. According to the nuclear hierarchical structure, binary masks of nucleuses, nucleolus and inclusions are created. For computing of topological characteristics, nuclei are presented as a hierarchy of binary images.

### 1 INTRODUCTION

Treatment of cancer is a worldwide problem (Yarbro et al. 2005). Unfortunately, the anticancer drugs that are presently used in the clinical practice have only limited success. To provide a significant clinical advance, new concepts have to be introduced to aid design of new tools for therapy. Understanding of mechanisms of tumor growth and progression is critical for development of novel anticancer drugs. Metastatic dissemination of cancer is primarily responsible for treatment failure, morbidity and death in cancer patients. Once cancer cells have spread and formed secondary masses (metastases), cancers are largely incurable despite the progress in medicine.

The main problem in understanding cell migration is that, as was recently shown, in vitro results concerning this phenomenon are of limited value in vivo. Thus, the only way to find out real data on cell migration is to study histological material obtained after surgery or biopsy. Colorectal cancer is a good model for studying tumor invasiveness in vivo (ex vivo) because of the easily histological detectable invasive front. Analysis of histology samples is one of the most important tasks of oncology diagnostics. A histological investigation is a final step before medical decision. There are many kinds of such investigations. They depend of painting histological samples. Usually such painting methodic connected with chemical representation of nuclear and nuclear is basic element for analysis on histological sample. Unfortunately body of nuclear has no connected pattern and include many false or no informative regions from anther nuclear, artifacts, cytoplasm painted ele-

ments and other noise. Therefore the task of real time nuclei shape description is very difficult (Figure 1).

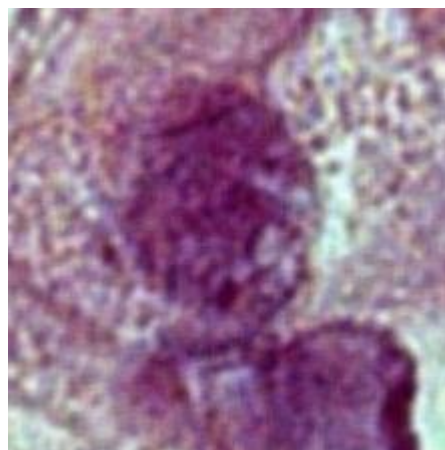


Figure 1. A fragment of histological image with nuclear.

A basic characteristic of nucleuses images is color of membrane and internal structure. Complexity of an image with nucleus is characterized by an internal topology their structure. Analysis of this structure is very important for pathology diagnoses.

The basic pathology process in cells starts from nucleuses. Therefore task of determination of nucleus pattern and its characteristic is important for analysis of pathology process in histology. This task can be divided into four sub-tasks:

- 1) isolation of nucleuses,
- 2) extraction of nucleus structure,
- 3) extraction of nucleus features such as size and density, grey level extrema, fractal dimension, texture parameters and shape measures, and

4) use of these features to classify the nucleuses (Orlov et al. 2008, Swedlow et al. 2003).

However, due to a complex, unstable image representation of histological samples, it is practically impossible to select or develop automatic methods that can be applied for solving these tasks. That is why most of the papers consider task of nucleuses image segmentation where only nucleus shape is extracted (Kim et al. 2007).

In this paper, algorithms for automatic computing of nucleus cell properties in histological images are presented. Shape, topological nuclear properties and nuclear orientation in tissue images are calculated that allow to determinate symptoms of cancer diseases.

It is very difficult to solve these tasks because they include many steps of analysis and processing. In the paper we present automatic methods for pre-processing and analysis of noisy nuclear images. The presence of noise is inevitable and is related to image acquisition procedure.

## 2 NUCLEUS PATTERN IN HISTOLOGICAL IMAGES

A nucleus is a membrane bounded structure that contains a cell and nuclear hereditary information and controls the cell growth and reproduction (Lodish et al. 2004). It is commonly the most prominent organelle in a cell. Nucleus is usually located inside a cell. Grey-level characteristics of nucleus differ from cell characteristics. In a nucleus, there is one or several rounded dark particles, which are named nucleolus.

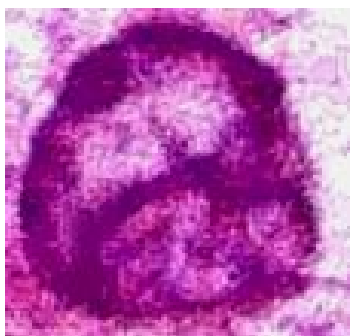


Figure 2. Image of a septa in nucleus.

The nucleus is bounded by a double membrane called a nuclear envelope. The structure of nucleus can be divided into membrane, septa (Figure. 2), inclusions, and nucleolus. Septa and inclusions constitute large structures that are crossing with membrane. Nucleolus constitutes internal structures without membrane crossing.

Structure of a histological image consists of several components: background, nuclear area, nucleus regions and artifacts. All the subpatterns have fuzzy borders in the digital image. A histogram analysis allows to separate background, nuclear and nucleus. But many of artifacts are presented in a whole range

of the image intensity, and this leads to errors in the segmentation results.

While artifacts prevent us to extract real contour of nucleus, false nucleus-like objects can be observed as well. In such cases results of a threshold segmentation have poor quality and false objects. Specialized segmentation algorithms have been developed to avoid the problems. They consist of two parts: nucleus extraction and nuclei detection.

## 3 NUCLEUS COMPONENT DETECTION

Nuclei components in an image are represented differently; each of them has individual characteristics and may be detected by different methods.

An image histogram analysis shows that the brightness histogram distribution has similar shape for all nucleus images. The background area is characterized by the gray values around the first (left) peak. This property leads to methods of background detection.

We use a watershed thresholding algorithm applied to the inverted histogram for the original image binarization, similar to (Sahoo et al. 1988).

The watershed transformation considers the gradient magnitude of an image histogram as a topographic profile. Bins having the highest gradient magnitude intensities correspond to watershed points, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed point flows downhill to a common local intensity minimum. Bins draining to a common minimum form a catch basin, which represents a segment.

Small areas (non nucleus particles) are removed by special procedure of area particles measurements (Figure 3.). This procedure is based on connected components analysis.



Figure 3. Result of nucleus detection by binarization of a histological image by watershed thresholding of the inverted histogram.

The algorithm of calculation of objects area was improved basing on neighborhood pixels analysis by

optimization of lines in stack. This algorithm support objects with any connectivity.

To scan the horizontal line is determined by input pixels. Its coordinates are calculated during process of line scanning and are write to the stack. It recursively performed an analysis of neighborhood and analysis of the lines. When the recursion ends, the stack from the extracted coordinates of the next line and analysis of lines repeated again. The algorithm works until the stack no lines. At the end of the raster scanning based on primary data collected from the lines, are the basic characteristics. Such organization of objects area calculation increase speed of processing.

After binarization the resulting image has poor quality in sense of extracted objects but it includes important information about nucleus border.

For definition of extracting method or algorithm image patterns of nucleuses can be divide to five classes by geometrical properties and positions:

- separate nucleus;
- low closing nucleus;
- middle closing nucleus;
- high closing nucleus;
- overlapping nucleus.

First class consists of nucleus with free border. This nucleus is no cross other objects. It is possible to process and analyze such objects after its binaryzation.

Low closing nucleus have common border with other objects. Fragments of common border should be less than nine pixels. In this case, separation is based on elementary morphological operation (Serra 1988 , Heng-Nian et al. 2007). It is simplest algorithm of separation. This algorithm consists of tree basic steps:

- erosion divide objects by decreasing their areas;
- identification mark objects by individual index;
- dilatation restore objects area and shape;

Such separation allows to serve characteristics of original nucleus.

Middle closing nucleus have border less 60%. In this case nucleuses have fragments for shape reconstruction. The watershed lines (Frucci et al. 2008) using is optimal algorithm for nucleus separation. There are many Algorithms for watershed lines construction, but result quality connects to distance metrics that are used in algorithms (Karaçalı et al. 2007, Bamford & Lovell 1998).

High closing nucleuses are the most difficult location for analysis. The borders of such objects are undefined. In result, these nucleuses are defined intuitively by interactive.

The case of overlapping nucleuses is extremely difficult. There is no clear solution to this problem. Therefore, for determining the characteristics of overlapping nucleus different approaches usually are used (Garrido, A. & Perez de la Blanca 200, Lu et al. 1998, Lu et al. 1998). These approaches connect to

determination a known shape and size of nucleuses. But more universal case of nuclear extraction is nucleus reconstruction from contour information.

Nuclear images have two main object characteristics for extraction: border and density of nucleus. Therefore a combination of methods is used here to define a border area: detection of nuclear's contour and detection by features.

#### 4 NUCLEAR CONTOUR DETECTION

Nuclear's contour is constructed from borders properties. Border should be relatively smoothed but artifacts are usually disrupt it. In order to solve this problem a special algorithm of border detection has been developed.

All small objects around nuclear are removed by skipping operation to provide further easy processing (figure 5.).



Figure 4 Simple thresholding of histology image with nuclear by inverse watershed method.

The morphological closing operation with very small structuring element removes small noise of binarisation. Then object's radii distribution of is collected. The radii are calculated along the lines from the center of mass to the border. This distribution includes many radii for every separate angle. For every angle all radii are removed except radii of minimal value. Then fitting line is constructed by spline smoothing (figure. 5).

For further correction current border result (figure 6) is united with border threshold segmentation (figure 4) by logical conjunction (figure 7).

Summarizing all mentioned above steps the algorithm of nuclear contour detection can be represented by scheme (figure 8).

Although we detect a smoothed nuclear contour without artifacts sometimes the constructed nuclear pattern is less than real object and few nucleus can be found outside of this pattern. In order to compensate

such effects an additional correction procedure by nucleus properties is necessary here.

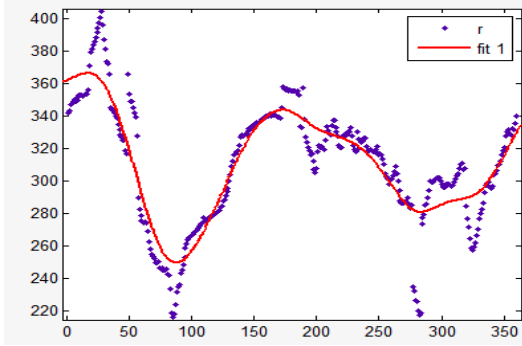


Figure 5 The distribution of radial radius from centre of mass to border ( $r$ ) and fitting this data by spline smoothing (fit 1).

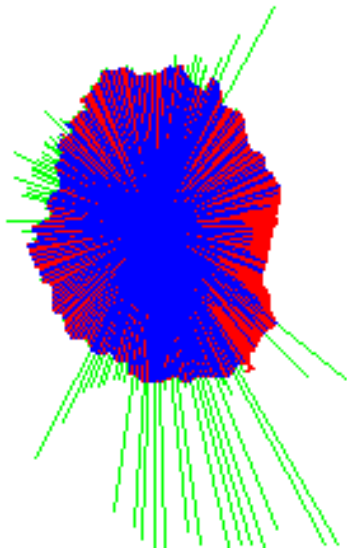


Figure 6 The Binary image of brain after border smoothing and distribution of minimal radius.



Figure 7 The result of consolidation threshold segmentation and border smoothing operation.

To obtain the topological characteristics, nucleus is presented as a hierarchy of binary images, where a binary image of nucleus is a root of graph, at the following level is the image of a net structure, the nucleolus is branches.

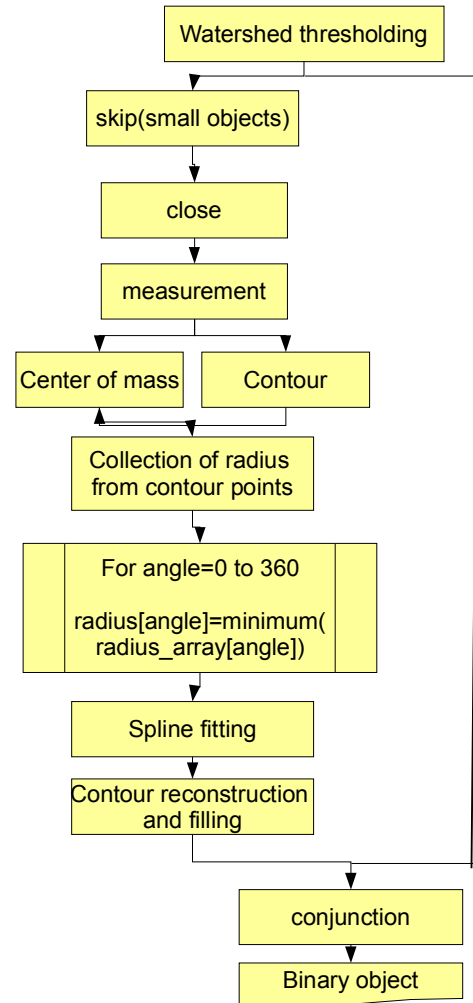


Figure 8 Algorithm for detection of smooth brain contour.

First of all, “opening” operation of mathematical morphology is carried out for separation of the merged close laying objects. Binary images of nucleus structures are consistently created. To obtain binary image of a nucleus, it is necessary to execute “filling” operation to area appropriate for a nucleus.

The following stage is necessary to obtain the image of nucleus large structures (septa, inclusions). At this stage the way of colouring of a histology preparation begins to play an important role. An image of nucleus large structure is extracted by means of masking the color image and additional threshold segmentation. For removal of noise and dust the morphology “close” operation is carried out. The objects with a size less than the established size of nucleus are removed. Then from this image, borders of nucleus are subtracted.

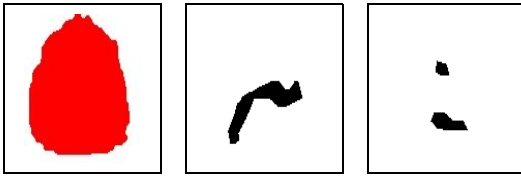


Figure 9 Binary images of nucleus, spike and nucleolus.

When binary images of nucleus, spike and nucleolus are determined (figure 9), they constitute multiphase hierarchical image (figure 10).

The idea of hierarchy of areas is easily realized in computer facilities using bits.

As each byte has eight bits, eight levels of hierarchy are optimum to allocate. Advantage of this method consists of sufficiency of one multiphase image used to draw conclusions on all levels of classification. (for example to get full information about nucleus).



Figure 10. Hierarchical image of nucleus.

Thus, pixel value of the nucleus in this point corresponds to 1. Pixel of nucleolus in a nucleus the pixel corresponds to 7. In case of presence of inclusion, which does not lay in a nucleus, value of pixel is 5. Therefore it is easy to operate with various nuclear components and then to analyze them

## 6 CONCLUSION

The algorithm of nuclear contour reconstruction was tested on image of histology samples with 100x magnification. Comparing of results of algorithms work show good quality (Table 1). The algorithm of contour reconstruction has best quality for many cases and allow to restore shape of nucleus in very complex situation when nucleus are overlapped.

Described method of nuclear segmentation detect as nuclear region are nuclear structure. It allows to modify result by removing the objects which are not belonging to nucleus and to construct the multiphase (hierarchical) image. This image reflects the hierarchy of nucleus structures. The obtained multiphase image of a nucleus can be used for subsequent analysis and measurements. Besides presence of nucleus and nucleolus at the multiphase image allows to study structure of a cell and its topological features.

Table 1. Quality of nucleus restoration by different algorithms

Objects/algorithm	Morphology segmentation	Water-shed segmentation	contour reconstruction
low closing nucleus	93%	86%	90%
middle closing nucleus	60%	92%	91%
high closing nucleus	0%	58%	79%
overlapping nucleus	0%	74%%	90%

Topology characteristics allow to describe features of histological tissue for cancer diagnostic (Ab-lameyko & Nedzved 2005).

## 7 ACKNOWLEDGEMENT

This work is partially supported by ISTC projects #B-1636 and #B-1682.

## 8 REFERENCES

- Yarbro, C-H. & Goodman, M. & Hansen Frogge, M. 2005. Cancer Nursing. *Jones & Bartlett Publishers*.
- Orlov, N. & Shamir, L. & Macura, T. J. 2008. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recognition Letters, vol. 29(11):1684-1693*
- Swedlow, J.R. & Goldberg, I. & Brauner, E. & Sorger, P.K. 2003. Informatics and quantitative analysis in biological imaging. *Science. Vol.300: 100-102.*
- Kim, K.-B. & Kim, S. & Kim G.-H. 2007 Nucleus. *Lecture Notes in Computer Science, Vol. 4432: 290-299.*
- Lodish, H. & Berk, A. & Matsudaira, P. & Kaiser, C.A. & Krieger, M. & Scott, M.P. & Zipursky, S.L. & Darnell J. 2004. *Molecular Cell Biology*. New York: WH Freeman.
- Sahoo, P.K. & Soltani, S. & Wong, A.K.C. & Chen, Y.C. 1988 A survey of thresholding techniques. *Computer Vision, Graphics, and Image processing, 1988, 41: 233-260.*
- Serra, J. 1988. *Image Analysis and Mathematical Morphology*. Vol. 2: Theoretical advances, Academic Press, Inc.
- Heng-Nian Qi & Feng-Nong Chen & Ling-Fei Ma. 2007. Pore Feature Segmentation Based on Mathematical Morphology. *Industrial Electronics Society, IECON 2007. 33rd Annual Conference of the IEEE Vol. 5-8: 2474 – 2477*
- Frucci, M. & Perner, P.G. & Sanniti di Baja. 2008. Case-Based Reasoning for Image Segmentation by Watershed Transformation. // *Case-Based Reasoning on Images and Signals. : 319-353*

Karaçalı, B. & Vamvakidou, A. P. & Tözeren, A. 2007. Automated recognition of cell phenotypes in histology images based on membrane- and nuclei-targeting biomarkers *BMC Medical Imaging*,:7-8.

Bamford, P. Lovell B. 1998. Unsupervised cell nucleus segmentation with active contours / *Signal Processing*, vol. 71 , Issue 2:203 – 213.

Garrido, A. & Perez de la Blanca, N. 2000. Applying deformable templates for cell image segmentation, *Pattern Recognit.*, 821-32, vol.33, no.5.

Lu, J. & Li, S. & Yang, J. & Liu, L. & He, A. & Liu, K. 1998. Lung cancer cell recognition based on multiple color space, *Proc. SPIE* vol.3522:378-86.

Kurugollu, F. & Sankur, B. 1998, Color cell image segmentation using pyramidal constraint satisfaction neural network, Proceedings of IAPR Workshop on Machine Vision Applications, 85-8, 1998.

Ablamejko, S. & Nedzved, A. 2005. Processing of optical images of cellular structures in medicine. Minsk.